Workshop Results – Next Steps

for the

LOS – Cost Estimation Team

Lead: Vanessa Griffin, ESDIS/SOO

Kathy Fontaine - GSFC

Bud Booth, Greg Hunolt, David Torealba – SGT

Claude Freaner - NASA HQ, Bruce Barkstrom - LaRC

Presentation to ESE
March 5, 2002

Levels of Service and Cost Estimation Study

Goals:

Develop baseline levels of service for data service providers:

Develop a cost estimation model that estimates life cycle costs for data service providers:

Approach:

- **Data service provider cost estimation must be grounded on an understanding of baseline levels of service and requirements.**
- **Develop a model that relates levels of service / requirements and model parameters (metrics) for a set of functional areas.**
- **Work with user community to establish baseline levels of service.**
- **Use Cost Estimating by Analogy as basic methodology.**
- **Build database of 'comparables' – existing ESE and other data activities.**
- **Develop cost estimating relationships through analysis of the comparables data base - test against independent cases, revise and improve model.**
- **Provide data service provider cost estimates, single and support architecture trade studies.**

Levels of Service Workshop

- The first SEEDS public workshop focused on establishing levels of service for SEEDS providers
- Goals:
  - Work with data service providers and end users to identify baseline levels of service needed from SEEDS era data service providers.
    - Focused on levels of service needed from peer data provider/users
    - Determine appropriate breakout of LOS by science discipline
  - Garner community support for proposed process leading to development of a SEEDS cost model and cost estimation tools.
    - Obtain feedback on Cost Team white paper which was distributed to attendees ahead of the workshop, sufficient to:
    - Refine the study approach, cost model parameters,

Key Results

- We had significant participation from the data provider community in considering levels of service in the SEEDS era.
  - We received 40 recommendations from the workshop that we are incorporating into our cost team planning
  - Progress made in developing the baseline levels of service.
- We also received critical feedback on our cost modeling approach.
- Many of the attendees submitted white papers that have led us to modify the cost and LOS study approach
- Additionally, all five study teams had the opportunity to share their status and discuss their efforts with the community.

Key Feedback from Workshop

- Cost model needs to emphasize flexibility in groupings of functionality into physical entities.
  - One size doesn't fit all – neither does seven
  - Consider use of pull down menus of functions to be "priced"
  - Logical differentiation between mission data centers, science data centers, and backbone data centers was not evident.
- Need to separate data search and ordering functions from distribution.
  - Likely that "search and order" service could be separated from the actual data management
- Need to consider need for highly educated, discipline expertise at the data service centers. Cost for these experts may not be included in commercial cost model
- Ensure that the cost model allows for subcontracting of functionality – perhaps to commercial firms

Key Feedback from Workshop (2)

- Several opportunities for getting additional feedback on SEEDS from user community were noted:
  - Take the message to the users. Talk to DAAC UWGs, ESIP advisory groups, etc.
  - Interview individual users – have DAACs and ESIPS recommend users to be contacted
  - Hold focus groups on LOS
  - Hold sessions on LOS at science conferences
- Consider using scenarios of what types of activities will be on-going in SEEDS, then look at appropriate approaches to meeting the scenarios, and ensure that cost model can provide estimates of costs for them.
- Reexamine approach for applications support.
- At the appropriate time, enlist the help of “tire-kickers” to independently evaluate the cost model.

Unfinished Work

- The short lead-time on the workshop announcement impacted the full participation of the user community
  - Most of the participants were data service providers.
  - We still need to hear feedback from the end-user community.
- We didn't make as much progress on establishing the levels of service as we had hoped.
  - Discussion primarily focused on high-level process issues rather than details of service.
  - Follow-up with the community, including additional feedback from end-users, is needed.

Workshop Recommendations

- In total, there were 40 recommendations arising out of the workshop discussions and white papers
- The cost model/LOS team is evaluating those recommendations and adjusting our plans accordingly.
- We have documented the summary of the recommendations, and our approach to incorporating them into our effort in a white paper, SEEDS Levels of Service – Cost Estimation Study: Synthesis of Workshop Results and Next Steps

Cost Modeling Issues

- Discussion both at and following the workshop have pointed out some concerns with using our cost modeling approach
  - Costing by analogy is considered by most experts to provide the best estimates of future costs.
    - However, Moshe Pniel has pointed out to the team that use of previous NASA missions to estimate costs for future ground systems may be misleading because of the NASA paradigm shift (i.e. smaller, better, faster)
    - For example, Pniel feels that the costs for the SeaWinds data systems were mostly driven by launch delays (with no concomitant ramping down of staff) and by a one size fits all approach.
  - Putting errors bars on cost estimates will be very difficult.
    - Experts at the workshop said the best we will be able to do is a factor of two – but will this be sufficient for NASA planning?

Cost Modeling Issues (2)

- Moshe Pniel has recommended that we consider a different approach to estimating costs for SEEDS components.
  - We need to look at the trends in costs for data centers over time and extrapolate the trends into the future
  - Moshe's example is the trends in costs for scatterometry data services – QuickSCAT data production and archive costs were lower than N-SCAT, while estimated costs for Cloudsat are lower still.
- The cost team is looking at how best to factor trends into our efforts.

Next Steps – Now through June Workshop

- February 19, 2002 – Begin site information collection effort; continues for a year or more.
- March 8, 2002 – Post Workshop Results, Next Steps, to SEEDS Website.
- March 31, 2002 – Complete update of levels of service per workshop results, cost model scenarios, etc., and post to SEEDS website.
- April and May 2002 – Get user feedback on updated levels of service.
- April 2002 – Build comparables database, adding Benchmark study data and new data as received; continues for a year or more.
- May 2002 – Develop initial set of cost estimating relationships (from comparables data) for prototype cost model.
- June 2002 – Demonstrate prototype cost estimation by analogy model at SEEDS workshop.

Back-Up Slides

Analysis of Workshop Results

- LOS – Cost Estimation Team reviewed input received:
  - Breakout Session Chairs' Notes
  - 14 White Papers submitted by Workshop Attendees
  - Team Member Notes
- From this input, the team derived a set of key results (next slide), and noted 39 recommendations.
- The team is reviewing the recommendations, intending to accept all (a few that seem problematic are under discussion), working out conflicts as needed.
- The team has developed a set of 'next steps':
  - Covering present through June Workshop
  - Actions from workshop and planned work.
- The team will post workshop results, including responses to white papers, and next steps to the SEEDS website ASAP.

Recommendations from Workshop

1. Emphasize that data service providers are abstract 'logical providers' or 'sets of functions' vs. physical entities. (Will do, March)
2. The cost estimation tool should offer providers a full menu of clearly defined and distinct functions to pick from, i.e. the general reference model rather than preset DSP type subsets. (Will do, reflect in March, prototype in June.)
3. Use scenarios to show how the cost model will be used, e.g. "HQ wants to fund missions, requires data management support, which means picking a set of functions, then look at alternative approaches to doing it, and estimate costs for them". (Will do, March.)
4. Outreach to Users:
    a. Take the LOS message to the users, e.g. DAAC UWG's (are there similar groups for ESIPs?), users of other types of existing systems (seen as a near term activity);
    b. Interview individual users; DAACs and ESIPS could suggest users, another idea was to ask workshop attendees to provide lists of users to be interviewed.
    c. Discuss LOS with Focus Groups – DAACs and ESIPs could suggest users to participate;
    d. Hold sessions on the LOS at conferences like AGU.

(Will do something - need a plan, need to coordinate with Formulation Team.)

5. When the time comes we get users to do independent tire-kicking of cost model. (Will do, future. Possibly informally with June prototype - TBD.)

Recommendations (2)

6. Make the model allow for ‘subcontracting’ of functions, e.g. archiving to backbone data centers, possibly commercial entities. (Will figure out a reasonable way - this is at heart of SEEDS approach, have to figure it out - post June.)
7. Small data centers might team to respond to an AO, providing more functionality together than they do on their own - recommend that cost model handle teams. (Same as above.)
8. Revamp the whole applications center area, starting with a better definition / description. (Will do, goal is March, will need review.) HQ Strategic Plan for applications available?
9. To do cost estimate, you have to understand ‘environment of use’, which may be different for different communities, and so different LOS may be appropriate for different communities. Recommend that model allow for this. (Will do, but post-June.)
10. Recommendation that the description of functional areas and requirements for the Backbone data centers be strengthened. (Will do, March.)
11. A user-oriented point of view of functions and levels of service, as exemplified in the [Kempler] white paper, should be adopted, at least for review by users, as being more directly relevant to their concerns.
12. Points in the [Kempler] white paper that expand on or add to the current description of functional areas, requirements, levels of service should be incorporated. (Will do, March.)

Recommendations (3)

13. The[Chen et al] paper recommends that the SEEDS initiative needs to take "these high level functions" (see section 5) into account explicitly in its planning, design, and cost estimation efforts." (Will do, March.)
14. The [Chen] white paper also contains twelve specific comments on the SEEDS working paper, addressing what the authors see as weaknesses in the current draft and model. Taken as a recommendation that the model be strengthened per the comments provided. (Will do, March.)
15. Taken from this [Wolf] paper for the LOS Cost Team activity is a recommendation that the description of user support be expanded to include points mentioned by the author. (Will do, March.)
16. Add Billing and Accounting, which was left out of functional breakdown. (Well, we will need this if ESE goes forward, but not at first - there will be no comparables, post-June.)
17. It [Collins white paper] provides a set of high level functional requirements ("musts") for National Data Centers. It is taken as a recommendation that the LOS - Cost Team's general data center reference model be examined for consistency with the paper, in its description of functional areas and levels of service / requirements. (Will do, March.)

Recommendations (4)

18. The [Kafatos] paper … suggests that SEEDS data service providers need to offer more on-line accessible services such as data subsetting, reformatting, but also analysis by user provided software, i.e. that providers should offer processing resources to users such that user software compliant with certain TBD standards could be transferred to the provider, executed on data available at the provider, with the results of the analysis returned to the user. This is taken as a recommendation that this functionality be called out as a level of service for SEEDS data service providers. (Will take it in hand as a feature for the post-June future, there are no comparables now.)
19. Call out search and access (order?) out (from distribution) as a separate functional area. (Will do, March.)
20. This [Jones] short paper consists of a recommendation that outreach, education, and training for the user community be included as an explicit activity under user support. (Will do, in a simple way, March.)
21. Lots of confusion over multi-mission data center type. Conflicts with mission data center type and other types. Recommend name change to include 'systematic measurements' or 'long term measurements'. (Will do, March.)
22. The LOS / Cost study should look at ESSP missions, which require data system support. (Taken to mean adding information on ESSP missions to database as possible, and eventually supporting estimation of ESSP data handling costs.)

(Will do as part of information collection effort, post-June.)

Recommendations (5)

23. Include cost of data acquisition, e.g. when an Applications Center must buy Landsat or other commercial data. (Will do - can add a line for this, so that a user of the model can include these costs as part of mission data, post June.)
24. Institutional Overhead rates must be included. (Will do - now included in labor rates, which are to be 'fully loaded', and provision for a charge per square foot for utilities, etc., March.)
25. The cost model should account for the fact that hardware costs fall with time (Will do, in some simple way - technology cost curves, a start by June.)
26. Especially where there is a significant cost impact, instead of assuming one level of service, the model should assume a distribution - i.e. some fraction at minimum level, some fraction at higher levels.) (Will do in a simple way downstream, post-June.)
27. Recommend that configuration management for science algorithms, which requires a significant effort, be called out as a cost factor. Note that product documentation should include processing history. (Will do in a simple way, March.)
28. Recommendation: Consider adding an interdisciplinary data center type, given the increasing importance of interdisciplinary science (noted by Jack Kaye). (Will do by making sure that interdisciplinary science is within scope of science data service provider type, March.)
29. Recommend that cost model recognize that multi-mission data centers need mission/instrument expertise for multiple missions - significant effort, a cost factor. (Will do in a simple way, June.)

Recommendations (6)

30. Make reprocessing schedule driven – then let rates and capacities follow from the schedule, and drive system sizing and cost. (Will do if possible, requires an approximate understanding of the schedule and processing requirements on a product by product basis which may not be practical, when TBD but post June probably.)
31. Recommendation: have one realistic level of service rather than multiple levels. Perception exists that as a provider you should always bid the minimum level of service since that's all the funding you'll get. The science community presses for high, well above minimum levels, but providers are never funded to provide them, ensuring a failure to meet expectations. (Need to discuss. AO/RFP could require response to a single (set of) levels, allow proposers to bid costed options for higher levels as they desire - would provide a common basis for evaluation, preserve option for pluses.)
32. Documentation (ranges from structural metadata and DIFs to full program documentation, ultimately to publication), recommended LOS:

Simplest - DIF, structural metadata

Next - ATBD (need to make sure these are as-built!)

Next - Guide or equivalent for users

Last - publication

(Will take this into account - whether explicit LOS as written TBD, March.)

Recommendations (7)

33. Include mission documentation – s/c instrument design, ICDs, etc., needed to describe how data was processed end to end. (Will do, March.)
34. Data coming to BBDC's from other DSPs … need to cover that on BBDC side. (ingest) If data is to be migrated to a BBDC for archive and distribution, from other DSPs, then cost to archive and distribute must be accounted for, BBDC's can't eat those costs. Recommendation: Cover BBDC support to other DSPs in model. (Will find a way to do, but not at first, post-June.)
35. Add 'packaging' to services list (for Bruce, but also cover in functional area / distribution). (Will do, March.)
36. Add data integration, which was suggested as a missing function. (Will include it under distribution as feature associated with a high level of service, when TBD, but post-June.)
37. Regarding Technical Coordination - seen as important, should be budgeted for, recommend that travel costs be included. Note that this is in reference to coordinating transfers of data between sites - add to scope of technical coordination. (Will do, small change but travel costs are grist for serious warfare, March.)
38. The cost model should consider costs end to end – should consider platform and instruments… we are starting with level 0 as input, but what happens upstream can affect costs below. (Need to discuss. We may be able to do something in the future, post-June)
39. Don't mention NARA requirement. May be a step back, may add unnecessary cost. (Will do, March.)

Recommendations (8)

40. Re user support LOS - distinguish between user types – some require more support. John Townshend suggested higher levels of service for key users/customers, i.e., gold/platinum card service. (Need to discuss.)

White Papers Received from Participants

Fourteen White Papers were written for the workshop.

1. **"User Oriented Services Model", Steve Kempler**
2. **"SEDAC Inputs to SEEDS Levels of Service Workshop", Bob Chen, Chris Lenhardt**
3. **"Operational User Support (OUS) Manifesto", Hank Wolf**
4. **"Distributed Data Access, Analysis, and Standards for Earth Science Data", Menas Kafatos**
5. **"Outreach, Education Training", Brenda Jones**
6. **"Data Management and Services for Global Change Research", Don Collins**
7. **"Data Services", Bruce Barkstrom**
8. **"SEEDS White Paper", Tom Kalvelage**
9. **"Input to be used in the SEEDS Public Workshop", Denny Thurman (EDC LTA)**
10. **"MODIS Land Rapid Response Activities…", Mark Carroll, et. al.**

☆ Key recommendations for cost model and LOS

White Papers Received (2)

11. **“FALCON: An Approach to Data System Interoperability for ESE Missions”, T. Yunck, et. al.**
12. **DAAC Alliance Response to NewDISS Formulation Questions”, DAAC Alliance Managers.**
13. **“Brainstorming on a 21st Century Earth Science Data System”, Skip Reber**
14. **“Comments on SEEDS Formulation Workshop”, Robert McGrath**